The Obama Effect: Patterns of Geographic Clustering in the 2004 and 2008 Presidential Elections

Nicholas R. Seabrook*
The Obama Effect: Patterns of Geographic Clustering in the 2004 and 2008 Presidential Elections*

Nicholas R. Seabrook

Abstract

This paper investigates patterns of geographic clustering in the county vote for the presidential elections of 2004 and 2008. Using local indicators of spatial association (LISA), the statistically significant areas of Democratic and Republican support are identified for both elections, thereby uncovering major sources of spatial autocorrelation in recent presidential voting. Most interestingly, the change in the presidential vote from 2004 to 2008 appears to be significantly more spatially dependent than the raw vote itself, suggesting that the so-called “Obama Effect” also exhibits a high degree of geographic clustering.

KEYWORDS: geographic clustering, presidential elections, Barack Obama

*Nicholas R. Seabrook is a doctoral candidate in the Department of Political Science at the University at Buffalo, SUNY.
The presidential election of 2008 was one of the most significant in U.S. history, as for the first time Americans elected an African-American candidate to the nation’s highest office. Despite predictions that widespread racism might produce a so-called “Bradley Effect,” whereby pre-election polls provide misleading levels of support for minority candidates due to systematic misreporting of voting intentions, Barack Obama’s performance in 2008 was very much in keeping with pre-election forecasts based on polling data in the vast majority of states. While race is undoubtedly a large part of the story of the 2008 election, which saw unprecedented levels of African-American turnout in many parts of the country, especially those with early voting (McDonald 2009), the substantive environment surrounding the presidential campaign was dominated by the economy. From what had been shaping up as an extremely close election, based on tracking polls following the party conventions in September, the Wall Street meltdown and subsequent economic crisis pushed Obama to a comfortable lead in the polls that would persist until Election Day, and more than any other factor is responsible for his election victory (Campbell 2009).

Those are the most familiar and widely acknowledged aspects of the election. This study instead investigates the degree to which spatial effects have been present in voting patterns in the two most recent presidential elections. While we know that Americans have become more politically polarized over the past few decades, there is some evidence to suggest that there has been a significant increase in geographic polarization as well, with Americans increasingly electing to live in more homogenous communities that reflect their own social and political values (Bishop 2008). If geographic polarization has become a defining feature of the U.S. electorate, we might expect to see significant spatial dependence and clustering both in presidential voting behavior, and in patterns of electoral change. In identifying areas of statistically significant clustering, this study also points to potential explanations for these effects, based on the varying impact of factors such as race, the economy, and evaluations of President Bush.

Data

The data used in this analysis consist of the 2004 and 2008 presidential election returns for each of the 3,257 counties and county-equivalent areas in the coterminous 48 U.S. states. These data were originally compiled by Dave Leip, election geographer, from information made available by the states themselves.

1 Alaska and Hawaii were excluded from the analysis, as the spatial weight matrix used to calculate subsequent spatial statistics in based upon contiguity.
through their elections boards and secretaries of state’s offices, and were attached to a shape-file of counties and county-equivalent areas from the 2000 Census using a Geographic Information System (GIS).

While the choice of the county as the unit of analysis for the study of the spatial dynamics of presidential elections is somewhat arbitrary, dictated here by convenience, given the availability of comprehensive county electoral data not accessible at other levels of geographical disaggregation, the county is nevertheless a politically relevant geographical subdivision in the United States. While individuals may not feel the same level of identification with their county as with more socially proximate aerial units such as the neighborhood or community, county borders often subdivide areas of some historical significance, and thus are contextually relevant environments to their inhabitants (Huckfeldt 1979). And while the county may not have the same political significance as city and legislative district boundaries, individuals nonetheless both pay county taxes and vote in county elections, also lending a sense of political identity to the county one resides in. The county is a particularly useful unit of study for geographical analyses in political science, given the wide range of economic and socio-demographic variables that are available at the county level through the Census (Cho and Gimpel 2009).

Geographic Clustering in the Presidential Vote

The first step towards identifying and explaining patterns of geographic clustering in the presidential vote by county is to examine visually the geographic distribution of support for the presidential candidates of the two major parties in 2004 and 2008, shown in Maps 1 & 2. Counties are color-coded according to the percentage of the two-party vote received by the Democratic presidential candidate in each election, Senator John Kerry of Massachusetts in 2004 and Senator Barack Obama of Illinois in 2008. Their Republican opponents were President George W. Bush and Senator John McCain of Arizona, respectively.

Red counties indicate areas of Republican strength, while blue areas represent areas of Democratic strength, with darker colors indicating a larger percentage of the two-party vote for that party’s candidate. Purple areas then indicate the most competitive counties in each election, those that fell within ±5% of parity.

Map 1. Democratic % of the Two-Party Presidential Vote by County, 2004

Map 2. Democratic % of the Two-Party Presidential Vote by County, 2008
In general, the geographic distribution of each party’s areas of support is fairly similar for the two elections, although the map for 2008 inevitably contains more blue and purple areas, reflecting the superior overall performance of the Democratic candidate in that election.

Areas of greatest Republican support appear to be clustered in the central region of the country, beginning in Texas and spreading up through Oklahoma, Kansas, Nebraska, and parts of eastern Colorado. Another Republican cluster is evident in the Mountain West, most notably in Idaho, Utah, northern Nevada, and eastern Oregon, spreading into parts of Montana, Wyoming, and the Dakotas. On the Democratic side, large clusters with high levels of Democratic support are evident in the New England states, in southern Texas, and along the California coast. In 2008, Democratic clusters also emerged in the upper Midwest, along the Southern Black Belt, and in New Mexico, spreading into parts of southern and central Colorado. Overall, Republican Party support appears to be clustered in larger areas with more dispersed populations, while Democratic support is concentrated in smaller areas with higher population densities.

The exact magnitude of the spatial dependence in the presidential vote at the global spatial scale can be ascertained through the calculation of Moran’s I statistic (Rogerson and Yamada 2009):

\[
I = \frac{m \sum_i \sum_j w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{(\sum_i \sum_j w_{ij})\sum_i (y_i - \bar{y})^2}
\]

Moran’s I is a weighted correlation coefficient that measures the extent to which the values of proximate observations \((i, j)\) are similar on a particular variable \((y)\) across a set of regions \((m)\). Figure 1 presents Moran’s I for the Democratic percentage of the county two-party vote, computed using a first-order queen-contiguity spatial-weight matrix \((w_{ij})\). The figure reveals the presence of strong and statistically significant positive spatial autocorrelation in the data for both election years, with the value of Moran’s I equaling 0.56 in 2004 and 0.58 in 2008.\(^4\)

---

\(^3\) This and all subsequent weight matrices and spatial statistics were calculated in GeoDa v.0.9.5-i (Anselin 2006), using Shape files obtained from the U.S. Census Bureau to which electoral data was attached in ArcGIS v.9.2 (ESRI). The first-order queen-contiguity matrix allows for the comparison of the value of a given variable in a county with the values of that same variable in all other contiguous counties.

\(^4\) The null hypothesis that the data were randomly distributed across the study area was tested using Monte Carlo simulation.
Figure 1. Univariate Moran Scatter Plots for Democratic Two-Party Vote %

2004

Moran's I = 0.5609

2008

Moran's I = 0.5834
The variables in the scatter plots are standardized so that the units correspond to standard deviations in the Democratic two-party vote percentage and its spatial lag, rather than the raw values. They reveal that the percentage of the two-party popular vote for the Democratic candidate in one district is strongly related the Democratic percentage of the vote in neighboring districts. Partisan strength is therefore non-randomly distributed across the electoral landscape in both the 2004 and 2008 elections, although the degree of spatial dependence is not significantly higher for 2008 than it was for 2004. While the spatial polarization of the electorate may well have been increasing in recent decades, this change has not been evident in the two most recent presidential elections.

A visual inspection of the electoral map for 2004 and 2008 appears to indicate a high degree of clustering in certain areas, and the value Moran’s I reveals that there is significant spatial autocorrelation in the data – also pointing towards geographic clustering of partisan support in the nation as whole. Yet neither of these approaches is sufficient for the statistical detection of areas of geographic clustering on a local spatial scale, or for determining which apparent areas of local clustering on the electoral map are statistically significant. One method which can be used to identify and distinguish areas of local clustering in the presidential vote is the calculation of local Moran statistics for each subregion in the dataset, also known as Local Indicators of Spatial Association (LISA) (Anselin 1995).

The local Moran statistic is an adaptation of the global Moran’s I statistic, and is used to assess the level of local spatial autocorrelation that exists around a particular point or subregion \(i\) (Rogerson and Yamada 2009):

\[
I_i = \frac{m(y_i - \bar{y})}{\sum_j (y_j - \bar{y})^2} \sum_{j} w_{ij} (y_j - \bar{y})
\]

This is equivalent to the global Moran’s I disaggregated by sub-region, so that the sum of the local Moran values obtained using the above formula will be equal to the global Moran’s I statistic multiplied by the sum of the weights \(w_{ij}\), in this case a first-order queen-contiguity spatial-weight matrix (Rogerson and Yamada 2009). The local Moran statistic essentially captures the extent to which the value of a particular variable \(y\) at a specified location \(i\) is similar or dissimilar to the values of that same variable at all surrounding locations \(j\).

By calculating the local Moran statistic for each sub-region, it is possible to create a map that displays the areas of local geographic clustering that are statistically significant, color-coded according to whether the cluster contains high values surrounded by other similar high values (dark red), low values surrounded by other similar low values (dark blue), or values surround by dissimilar values,
such as high surrounded by low (light red) or low surrounded by high (light blue). White areas of the map indicate that the local Moran statistic falls short of statistical significance.\textsuperscript{5}

The local Moran statistics for the 2004 and 2008 county presidential vote are illustrated in Maps 3 and 4.\textsuperscript{6} A number of areas of clustered partisan support are evident in the maps for both elections. The largest Republican clusters are once again apparent in the central region of the country, from Texas up through Nebraska, and in the states of the Mountain West. Smaller clusters of Republican support are also scattered throughout the South and in a few rural Midwestern areas.

Map 3. Geographic Clustering in Two-Party Presidential Vote by County, 2004

---

\textsuperscript{5} The significance level for each local Moran statistic was determined through Monte Carlo simulation of the null hypothesis.

\textsuperscript{6} Due to the default color scheme in the GeoDa software package, which assigns red to clusters of high values and blue to clusters of low values, the local Moran statistics were calculated using the Republican percentage of the two-party vote, thus resulting in red coloring for clusters of Republicans support and blue coloring for clusters of Democratic support.
Democratic support manifests itself in a larger number of generally smaller clusters than Republican support. These Democratic clusters tend to be centered on densely populated urban areas, in contrast to Republican clustering, which is generally more evident in rural sections of the country. Several statistically significant clusters of Democratic support can be attributed to the party’s urban strength: one in the New England states, spreading down into the New York and Philadelphia metro areas; another in the Washington D.C./Baltimore metro area; a third in eastern Ohio, spreading from Cleveland down into western Pennsylvania and encompassing the Pittsburgh metro area; plus four more centering on Detroit, St Louis, the San Francisco Bay, and the Seattle/Portland corridor.

Other Democratic clusters reflect the party’s strong appeal to Latino voters: one in Southern Texas and another in northern New Mexico; its strong appeal to African-American voters, scattered across the Southern Black Belt, with additional clusters that seem to center on the Mississippi River; plus one in the upper Midwest for parts of Minnesota, Wisconsin, Iowa, and Illinois, and another in the deep south for parts of Louisiana, Mississippi, and Arkansas. A further cluster can be identified in northern Colorado. Overall, then, there is evidence of a high degree of local geographic clustering in both presidential elections.

The LISA cluster maps also provide insight into how the areas of statistically significant clustering in the presidential vote have changed from 2004.
to 2008, greater insight than can be gained through a simple visual inspection of
the raw-vote percentages. From a comparison of the two maps, the large
Republican cluster in the center of the country appears to grow from 2004 to
2008, spreading in an easterly direction into parts of Arkansas, Louisiana, and
Missouri, whereas the large Mountain West cluster appears to shrink, most
notably in Montana, Oregon, and Nevada. There is also a fairly substantial
increase in Republican clustering in the South in 2008, especially along the Gulf
Coast and in rural sections of Tennessee, Kentucky, Alabama, and Georgia.

Interesting changes can also be seen in the clustering patterns of
Democratic support, with the Democratic clusters in New England, along the
Southern portion of the Mississippi River, and in eastern Ohio/western
Pennsylvania—each shrinking significantly from 2004 to 2008. In contrast, those
Democratic clusters in the upper Midwest, the Northwest, northern California,
and New Mexico each experienced growth, with Democratic support spreading
into additional surrounding counties. There is actually evidence of a new
Democratic cluster appearing in the Dakotas. Conversely, patterns of Democratic
support in the Southern Black Belt region are remarkably similar in the two
elections, despite the presence of an African-American candidate on the ballot in
2008.

The evidence presented in this section suggests that there was a high
degree of geographical clustering at both the global and local scale in the
presidential vote by county for each of the two most recent presidential elections,
with the two major parties drawing their support from very different regions of
the country. Despite this, the overall level of geographic polarization in the
electorate has not significantly increased in the last 4 years, perhaps indicating
that a plateau was reached somewhere between 2000 and 2004, with increased
polarization in some areas now being offset by growing spatial heterogeneity in
others.

It is also interesting to note that many of the geographic clusters of
partisan support evident in the county two-party vote reach across state lines,
sometimes encompassing all or part of several (or sometimes many) different
states. While presidential election results are aggregated to the state level by
constitutional design, it seems clear that the Electoral College map obscures many
significant local and regional voting patterns that do not confine themselves
within the boundaries of individual states. The next section focuses on the change
in the county presidential vote from 2004 to 2008, in order to isolate the extent to
which spatial effects either contributed to or perhaps even hindered the election of
Barack Obama.
The Obama Effect: Change in the County Vote from 2004 to 2008

In order to examine patterns of clustering and spatial dependence in the change in partisan support from 2004 to 2008, the difference in the Democratic and Republican percentages of the two-party vote between the two elections was first calculated for each county, in order to allow for visual examination of the geographic distribution of electoral change (shown in Map 5). Unsurprisingly, given that Barack Obama received 53.69% of the two-party vote in 2008, compared to John Kerry’s 48.76% in 2004, the majority of the map is blue, representing a swing towards the Democratic Party from 2004 to 2008. The most dramatic increases in the Democratic vote percentage, those of 10% or more, can be seen in the state of Indiana, in parts of Vermont, in southern and western Texas, and in parts of North Dakota and Montana.

Map 5. Change in the Two-Party Presidential Vote from 2004-2008 by County

Perhaps of greater interest is the relatively small number of counties where John McCain was actually able to improve on the performance of President George W. Bush, despite his smaller share of the popular vote overall. Some of these effects are surely idiosyncratic, such as the Republican gains in parts of Arizona and Massachusetts, which may reflect nothing more than the presence on
the ballot of Arizona Senator John McCain in 2008 and the absence of Massachusetts Senator John Kerry, the Democratic nominee in 2004. Other areas, however, seem to indicate a more systematic effect, most notably the large band of counties with improved Republican performance stretching from Texas across the Southern U.S. and into Appalachia. Beginning at the extreme northern edge of this band, the Republicans improved on their 2004 vote percentage in parts of western Pennsylvania and eastern Ohio, large parts of West Virginia and Kentucky, a small area of western Virginia, almost all of Tennessee, Arkansas, Oklahoma, and Louisiana, and parts of northern Alabama, Mississippi, and Georgia.

The presence of such a large and geographically concentrated region of increased Republican support is remarkable given their declining vote share from 2004 to 2008, and indicates at least preliminarily that the forces driving electoral change in this most recent election cycle have a significant spatial component to them. Moreover, by controlling for the national swing in the presidential vote from 2004 to 2008, Map 6 allows this apparent cluster to become even more pronounced.

Map 6. Change in Presidential Vote Standardized by National Swing by County
What stands out when manipulated in this way is a much more substantial and geographically distinct region of the country, where McCain was able to outperform expectations, given his overall vote.\textsuperscript{7}

Figure 2 adds weight to this conclusion, by illustrating the global Moran’s I statistic for the change in the two-party presidential vote by county from 2004 to 2008, again using a first-order queen-contiguity spatial-weight matrix. While the raw-vote percentages themselves exhibit a high degree of spatial dependence, the level of spatial autocorrelation in the change in the vote percentage is an even higher 0.73, indicating that across the electoral map, changes in one county were strongly associated with changes in surrounding counties. In trying to isolate the “Obama effect”, therefore, we should be extremely cognizant of factors that might vary systematically in their impact in different areas and regions of the country, and thus might account for the dramatic spatial effects evident in Figure 2.

\textbf{Figure 2.} Univariate Moran Scatter Plot for the Change in Democratic Vote \%

\textsuperscript{7} Map 6 was generated by adjusting the 2004 county two-party vote percentage based upon an assumption of universal national swing across all counties, and then calculating the change from 2004 to 2008 based upon this adjusted expectation.
Map 7. Geographic Clustering in the Change in the Presidential Vote by County

Map 7 illustrates the local Moran statistics for the change in the presidential vote from 2004 to 2008, allowing us to identify the areas of statistically significant clustering in the data. The patterns of geographic clustering in the change in the two-party vote are dramatically different from the patterns observed in the raw-vote percentages, suggesting that there are very different spatial dynamics driving partisan electoral change from those associated with underlying levels of party strength. The large cluster of improved Republican performance in the South once again stands out on the LISA map, encompassing several distinct interconnected bands of counties. A notable area of clustering is evident in the existing Republican strongholds of northern Texas and western Oklahoma, corresponding closely to the clustering in the raw vote totals for 2008 on Map 4.

A similar pattern emerges with the cluster located in areas of eastern Texas and into Louisiana, and is also indicative of Republican gains in areas of existing electoral strength. Interestingly, this cluster extends into the heavily white areas of northern Texas, but does not penetrate the Dallas or Houston metro areas, cities with African-American populations of 23.8% and 25.3% (U.S. Census Bureau), or the heavily Latino areas in the south and west of the state. The cluster also contains most of the state of Louisiana, other than those heavily African-American counties in the vicinity of the city of New Orleans.
Another cluster of improved Republican performance corresponds with areas of Republican clustering in the raw vote percentages in southern Kentucky, eastern Tennessee, northern Mississippi, Alabama and Georgia, and small parts of western Virginia and North Carolina. However, the clustering evident on the change map covers a much larger area of Appalachia than that on the raw vote percentage map for 2008, containing almost the entire states of Tennessee and West Virginia, and spreading into parts of Pennsylvania and Ohio. In other parts of the South, the Republican clustering in the change in the presidential vote is located in areas of strong Democratic support in the raw vote percentages, most notably in the states of Mississippi and Arkansas, in a band spanning the Mississippi river. These areas remained heavily Democratic in 2008 despite the significant swing towards McCain evident on the change map. Smaller Republican clusters are also evident in Arizona, Kansas, the Florida panhandle, and southern Georgia, the latter two of which are also consistent with clusters of Republican support on Map 4.

There is also evidence on Map 7 of statistically significant clustering in areas of improved Democratic performance in 2008. A large cluster emerges along the Mexican border, spreading from southern Texas through central New Mexico, which appears to correspond to areas with large Latino populations. While this cluster includes parts of southern Texas and northern New Mexico, where there is also clustering evident on Map 4, it also spreads into southern New Mexico and western Texas, areas containing mostly Republican-leaning or swing counties, even with the strong Democratic gains in 2008. Another cluster contains much of the state of California and parts of Nevada, and is also significantly larger than the cluster that emerges on the raw vote percentages map, where it is confined to the coastal area of northern California surrounding San Francisco. There are also several Democratic clusters on the change map in areas of the country that we might associate with a more libertarian set of political values: one in Montana, another in the Dakotas, and a third in New Hampshire and Vermont.

The change patterns in the state of Colorado are particularly interesting, as the significant Democratic gains came not from their existing strongholds around the cities of Denver and Boulder, but from Republican areas in the northern part of the state, where conversion of red counties to swing counties produced a cluster that helped propel Barack Obama to victory in that state. In northern Utah, the counties surrounding Salt Lake City remain part of a large cluster of Republican support on the 2008 raw vote percentage map, despite experiencing a significant swing towards the Democrats as illustrated by the statistically significant cluster on the change map. Urban clusters centering on Atlanta, Jackson MS, and Raleigh-Durham NC, also stand out, as does the metro-Atlantic cluster spreading from northern Virginia, through the Washington, D.C. and Baltimore metro areas, into Philadelphia. These latter two clusters also contain parts of red states that
Obama was able to capture in 2008, with the growth of the D.C. cluster into the northern part of Virginia contributing to his victory in that state, and his win in North Carolina in part made possible by Democratic gains in the “Research Triangle” area.

Another interesting geographic cluster of increased Democratic strength is located in the Midwest, centered on Lake Michigan, and containing parts of eastern Wisconsin, northern Illinois, western Michigan, and Ohio. This cluster also contains almost the entire state of Indiana, which saw a dramatic swing towards the Democrats in 2008, and given pre-election polling was perhaps the most surprising red-state to blue-state conversion in the 2008 election. Finally, there is a cluster on the change map located in western Iowa and eastern Nebraska, containing the cities of Omaha and Lincoln. In Nebraska, these large Democratic swings occurred in counties that remained solidly Republican in the 2008 election, though many parts of western Iowa were converted to swing counties, thus contributing to an Obama win in that state.

From the evidence in this section, it appears that many of the electoral effects that propelled Barack Obama to victory in the 2008 presidential election were heavily spatial in nature. While Obama lost support compared to the 2004 baseline in many heavily white Southern areas, this was more than offset by the impressive turnout of African Americans throughout the Black Belt, and his appeal to Latino voters in the Southwest. Most analysts agree that Barack Obama’s victory can be attributed more than anything else to the economy, and those geographic areas most affected by the collapse of the financial markets and subsequent economic slowdown, which occurred on the watch of his Republican predecessor, do appear to correspond closely to the areas where we see the most impressive Democratic gains in 2008.

This study has focused on identifying and comparing those areas of statistically significant geographic clustering illustrated in Maps 3, 4, and 7. Further research is clearly needed to ascertain the exact nature of these spatial effects, and to identify those independent variables that might help explain and predict the patterns of geographic clustering in the presidential vote. While this study represents a first step towards an improved understanding of spatial clustering in presidential elections, it has focused on only the two most recent electoral contests. Future research should also attempt to extend this analysis back in time, to ascertain how geographic polarization of the electorate and patterns of spatial homogeneity and heterogeneity in presidential voting have changed over time.
References


